Simultaneous EEG-fMRI brain signatures of auditory cue utilization

Mathias Scharinger[1]*, Björn Herrmann[1], Till Nierhaus[2] and Jonas Obleser[1]

[1] Max Planck Research Group "Auditory Cognition," Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany
[2] Department of Neurology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

Optimal utilization of acoustic cues during auditory categorization is a vital skill, particularly when informative cues become occluded or degraded. Consequently, the acoustic environment requires flexible choosing and switching amongst available cues. The present study targets the brain functions underlying such changes in cue utilization. Participants performed a categorization task with immediate feedback on acoustic stimuli from two categories that varied in duration and spectral properties, while we simultaneously recorded Blood Oxygenation Level Dependent (BOLD) responses in fMRI and electroencephalograms (EEGs). In the first half of the experiment, categories could be best discriminated by spectral properties. Halfway through the experiment, spectral degradation rendered the stimulus duration the more informative cue. Behaviorally, degradation decreased the likelihood of utilizing spectral cues. Spectrally degrading the acoustic signal led to increased alpha power compared to nondegraded stimuli. The EEG-informed fMRI analyses revealed that alpha power correlated with BOLD changes in inferior parietal cortex and right posterior superior temporal gyrus (including planum temporale). In both areas, spectral degradation led to a weaker coupling of BOLD response to behavioral utilization of the spectral cue. These data provide converging evidence from behavioral modeling, electrophysiology, and hemodynamics that (a) increased alpha power mediates the inhibition of uninformative (here spectral) stimulus features, and that (b) the parietal attention network supports optimal cue utilization in auditory categorization. The results highlight the complex cortical processing of auditory categorization under realistic listening challenges.

Keywords: audition, categorization, cue weighting, spectro-temporal information, alpha suppression, attention

INTRODUCTION

The interpretation of acoustic signals is an essential human skill for goal-directed behavior and vocal communication. The core process underlying this skill—auditory categorization—has been shown to be highly flexible and adaptive, and allows, for instance, speaker recognition in a cocktail party situation (Zion Golumbic et al., 2013), or speech comprehension in noise (Nahum et al., 2008). In both cases, attention has to be directed to the most informative aspect of the acoustic signal (Hill and Miller, 2010).

Neurophysiological studies have suggested that the relative weighting of information during categorization (*information gain* or *cue weighting*, cf. Holt and Lotto, 2006) may be subserved by the interplay between excitatory and inhibitory mechanisms (Thut et al., 2006; Rihs et al., 2007; Weissman et al., 2009). One promising neurophysiological marker of functional inhibition processes are brain oscillations recorded using electroencephalography (EEG), predominantly in the alpha frequency range (8–13 Hz, Foxe et al., 1998; Foxe and Snyder, 2011; Weisz et al., 2011, 2013; Klimesch, 2012). Initially, alpha power had been interpreted as reflecting the degree to which primary cortical areas are in an "idling" mode (Adrian and Matthews, 1934; Niedermeyer and Silva, 2005). More recent studies on auditory

comprehension, on the other hand, have shown that the processing of degraded speech stimuli is accompanied by relative decreases in alpha power suppression, i.e., relative increases in alpha power (Obleser and Weisz, 2012; Becker et al., 2013). One interpretation of this finding is that relative increases in alpha power index greater attention and working memory demands under degradation (Ronnberg et al., 2008; Wild et al., 2012). It has been further proposed that brain regions showing high alpha power undergo inhibition, which in turn allows enhanced processing of task-relevant information (Klimesch et al., 2007).

Brain areas underlying the processing and categorization of acoustic information have been identified by means of functional magnetic resonance imaging (fMRI). Previous studies have shown that the posterior part of the superior temporal gyrus (pSTG) is crucially involved in auditory categorization and discrimination (Hall et al., 2002; Guenther et al., 2004; Husain et al., 2006; Desai et al., 2008; Bermudez et al., 2009; Sharda and Singh, 2012). Importantly, in most of these studies, auditory categorization was also subserved by the planum temporale (PT) in the pSTG. The PT has recently received particular attention, because it does not only play a general role in auditory categorization (Griffiths and Warren, 2002; Husain et al., 2006; Obleser and Eisner, 2009) but also a more specific one with regard to the

processing of spectral information and pitch (Hall and Plack, 2009; Alho et al., 2014).

Furthermore, feature-selective attentional processes play a crucial role in categorization. Studies concerned with aspects of selective attention during categorization have mainly focused on the visual system (Yantis, 1993; Posner and Dehaene, 1994; Corbetta et al., 2000; Yantis, 2008). These studies identified the inferior parietal lobule (IPL) as an important, hub-like structure, being involved when participants focus attention on informative stimulus features (Shaywitz et al., 2001; Behrmann et al., 2004; Geng and Mangun, 2009; Salmi et al., 2009; Schultz and Lennert, 2009; Gillebert et al., 2012). Existing research on attention in audition has further provided evidence for the involvement of the parietal network (Rinne et al., 2007; Salmi et al., 2009; Hill and Miller, 2010; Henry et al., 2013). In addition, a recent structural imaging (voxel-based morphometry) study also highlighted the role of the IPL in categorization processes (Scharinger et al., 2014).

More recently, the possibility to combine recordings of EEG oscillatory activity and fMRI Blood Oxygenation Level Dependent (BOLD) activity has been explored in several imaging studies. Simultaneous EEG–fMRI recordings (Ritter and Villringer, 2006; Sadaghiani et al., 2010, 2012) suggest that alpha power can be negatively (Goldman et al., 2002; Laufs et al., 2003; Ritter and Villringer, 2006) or positively (Moosmann et al., 2003; Liu et al., 2012) correlated with brain metabolism, depending on the brain regions these correlations are observed in. However, multi-modal neuroimaging evidence on auditory cue weighting during categorization has been essentially absent. Most studies concerned with a functional coupling of alpha power and BOLD signal in selective attention tasks compared the processing of task-relevant information with the processing of task-irrelevant distractor information (e.g., Scheeringa et al., 2012).

It is thus less clear how multiple, potentially competing cues provided by the same acoustic stimulus, will be reflected in alpha-tuned functional processes and concomitant BOLD change. To this end, we designed two stimulus sets for auditory categorization. In the first stimulus set, categorization could be based on spectral properties or physical duration, with spectral properties being more informative. In the second stimulus set, sound duration became the more informative cue, while spectral properties could still be used for categorization. Using combined EEG/fMRI, we asked (a) whether auditory categorization yields a behavioral preference for the most informative stimulus cue in each condition; (b) which brain areas support change in cue utilization, (c) whether alpha power shows relative increases under degradation and (d) whether alpha power correlates with BOLD in brain areas dedicated to the processing of acoustic cues.

MATERIALS AND METHODS

PARTICIPANTS

Sixteen healthy volunteers were recruited from the participant database of the Max Planck Institute for Human Cognitive and Brain Sciences (7 females, age range 20–29 years, age 25 ± 2.7 years mean $\pm$ standard deviation). They were all right-handed, native speakers of German with no self-reported hearing impairments or neurological disorders. Due to technical problems with

EEG acquisition in the magnetic resonance (MR) scanner, we had to exclude one participant from further analyses. Participants gave written informed consent and received financial compensation for their participation. All procedures followed the guidelines of the local ethics committee (University of Leipzig) and were in accordance with the Declaration of Helsinki.

STIMULI

Stimuli were based on spectral and durational modifications of an inharmonic base signal. This base signal was constructed by adding 16 exponentially spaced sinusoids (ratio between successive components: 1.15) to the lowest sinusoid component frequency of 500 Hz (Goudbeek et al., 2009; Scharinger et al., 2014). We modified the spectral properties of individual sounds by applying a band-pass filter with a single frequency peak, using a second order infinite impulse response (IIR) filter with a bandwidth corresponding to a fifth of its frequency peak. The term "spectral peak" is henceforth used to refer to the filters' center frequency, which also describes the resulting spectral properties. Duration modifications were based on differences in the length of the sounds.

Individual members of category distributions, arbitrarily labeled "A" and "B," varied on the basis of spectral peak and duration: For individual sounds of each category, spectral filter frequencies and durations were randomly drawn from bivariate normal distributions. These distributions, with equal standard deviations, σ, differed in their means, μ, between the two categories, A and B (**Table 1**). Thus, each individual sound was characterized by the two dimensions, duration and spectral peak, with means of duration and spectral peak differing between the two category distributions. Each category distribution consisted of 1000 sound exemplars from which a random sample was drawn for each participant in the experiment. Following Smits et al. (2006), we converted spectral peak frequency and duration to scales that allowed for psychoacoustic comparability. Consequently, frequencies were converted to the equivalent rectangular bandwidth (ERB) scale that approximates the bandwidths of the auditory filters in human hearing (Glasberg and Moore, 1990), and durations were converted to a logarithmic scale (DUR; cf. Smits et al., 2006). **Table 1** illustrates the means (spectral peak and durations) of the category distributions in psychophysical and physical units.

In the first half of the experiment (*nondegraded condition*), the two stimulus distributions did not overlap in their spectral peak, but $\frac{1}{3}$ of the sounds in category A and B overlapped in duration (**Figure 1A** top). This set-up aimed at biasing participants to focus on spectral cues while sound duration may serve as secondary cue. In the second half of the experiment (*degraded condition*), spectral cues were modified by applying four-band noise vocoding to the original stimulus distributions (Drullman et al., 1994; Shannon et al., 1995). Noise vocoding was done by dividing the original signal into four frequency bands, extracting the amplitude envelope from each band and reapplying it to bandpass-filtered noise carriers with matched cut-off frequencies. Envelopes were extracted using a zero-phase, 4th-order Butterworth low-pass filter; the low-pass filter cutoff was set at 256 Hz. Scaling for equal root mean square (RMS) energy

Table 1 | Means and standard deviations (in parentheses) of spectral peak and duration distributions for stimulus categories A and B in the nondegraded and degraded conditions (psychophysical and physical units).

Stimulus category	Nondegraded		Degraded	
	A	B	A	B
Spectral peak (ERB)	20.00 (0.31)	17.00 (0.31)	16.80 (0.31)	15.50 (0.31)
Spectral peak (Hz)	1739 (8)	1196 (8)	1166 (8)	984 (8)
Duration (DUR)	47.70 (1.31)	52.53 (1.31)	47.70 (1.31)	52.53 (1.31)
Duration (ms)	118 (1.14)	191 (1.14)	118 (1.14)	191 (1.14)

was performed channel-wise for each channel envelope (Rosen et al., 1999; Erb et al., 2012). We chose four-band noise vocoding because it offers a well-established reduction of spectrally-based intelligibility (cf. Scott et al., 2006; Obleser and Kotz, 2010; Obleser et al., 2012), thereby ensuring comparability to studies on alpha power suppression in speech, while simultaneously being an ecologically valid modification by simulating effects of cochlear implants (Poissant et al., 2006).
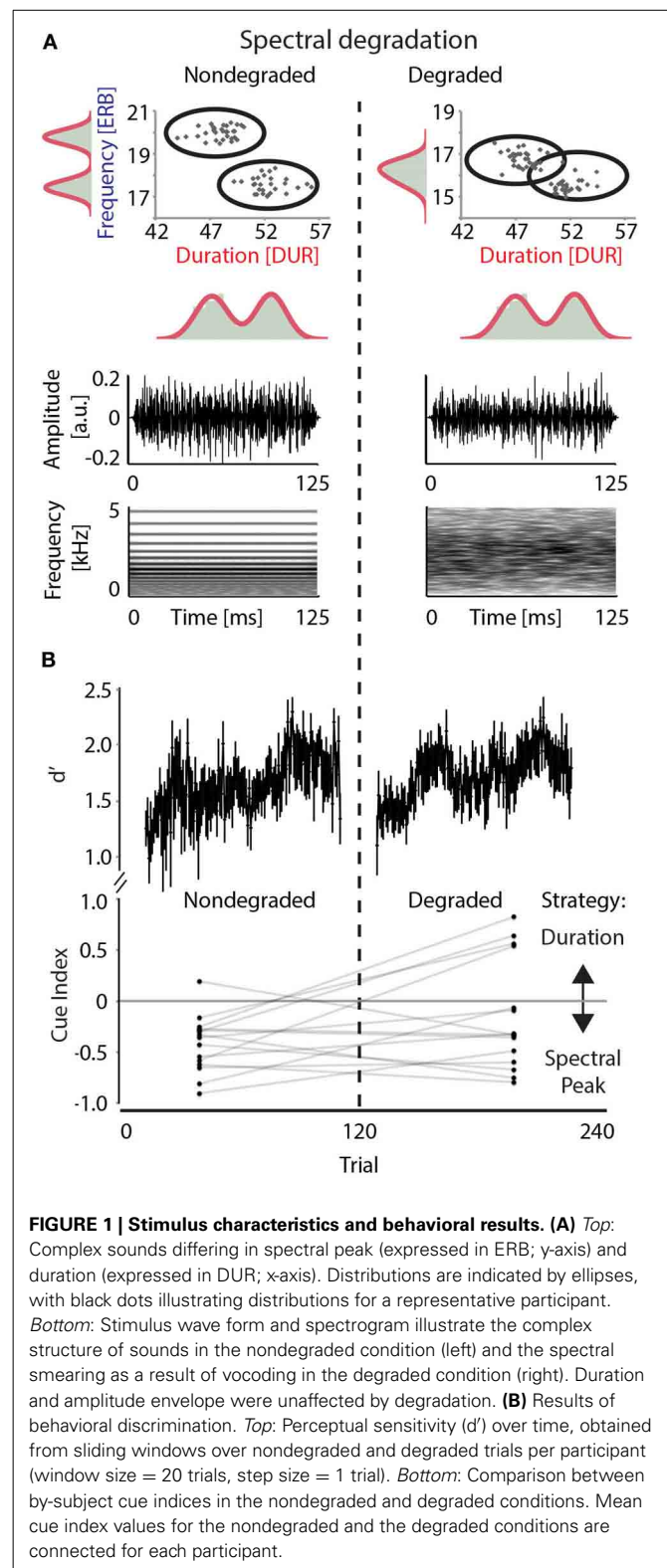
Noise vocoding led to a smearing of spectral detail, while amplitude envelope features and original stimulus duration remained unaffected (**Figure 1A**, bottom). Thus, as demonstrated before (Scharinger et al., 2014), we aimed at inducing a change in acoustic cue utilization, from spectral peak in the first (nondegraded) condition, to stimulus duration in the second (degraded) condition of the experiment. The stimulus degradation in the second half of the experiment therefore targeted the spectral properties (i.e., spectral peak, but also affected other spectral features such as harmonicity). Thus, degradation of the initially informative spectral cue ought to decrease participants' reliance on that cue and prompt a relatively increased reliance on the duration cue.

All stimuli were normalized for equal root-mean-square intensity and presented at ~60 dB SPL. Onset and offset ramps (5 ms) ensured that acoustic artifacts were minimized.

EXPERIMENTAL PROCEDURE

Participants were first familiarized with the categorization task in the scanner and had to complete a short practice run consisting of 20 sounds (10 from category A and 10 from category B) that did not occur in the main experiment. The subsequent main experiment was arranged in four runs: Two initial runs with nondegraded sounds, and two subsequent runs with spectrally degraded sounds (**Figure 1A**, top). In each run, 60 sound exemplars, randomly drawn from categories A and B with equal probability, were presented in a sparse imaging design in the MR scanner (Hall et al., 1999). The sparse design was chosen in order to guarantee that stimuli could be presented during silent periods in-between the acquisition of echo-planar images (EPI). At the same time, this design reduced contamination of the EEG signal by gradient switches during volume acquisition.

On each trial, one acoustic stimulus was presented on average 2 s after the offset of a preceding EPI sequence (±500 ms). Subsequently, a visual response prompt (green traffic light) was

FIGURE 1 | Stimulus characteristics and behavioral results. (A) *Top*: Complex sounds differing in spectral peak (expressed in ERB; y-axis) and duration (expressed in DUR; x-axis). Distributions are indicated by ellipses, with black dots illustrating distributions for a representative participant. *Bottom*: Stimulus wave form and spectrogram illustrate the complex structure of sounds in the nondegraded condition (left) and the spectral smearing as a result of vocoding in the degraded condition (right). Duration and amplitude envelope were unaffected by degradation. **(B)** Results of behavioral discrimination. *Top*: Perceptual sensitivity (d') over time, obtained from sliding windows over nondegraded and degraded trials per participant (window size = 20 trials, step size = 1 trial). *Bottom*: Comparison between by-subject cue indices in the nondegraded and degraded conditions. Mean cue index values for the nondegraded and the degraded conditions are connected for each participant.

presented on a screen which participants viewed through a mirror 3 s after stimulus onset. Participants were then required to indicate whether the presented sound belonged to category A or category B by pressing one of two keys on a button

box. Button assignment was counterbalanced across participants. Following the response, participants received corrective feedback (*Correct/Incorrect*), which was displayed for 1 s in the middle of the screen. Five seconds after the onset of an acoustic stimulus, a subsequent EPI volume (acquisition time TA = 2 s) was acquired, such that the BOLD peak would best capture stimulus processing. At random positions within each run, 15 silent trials (=20% of all trials) without required responses served as baseline. The duration of the entire experiment with short breaks between runs was 50 min.

ACQUISITION AND PRE-PROCESSING OF EEG DATA

The continuous EEG was recorded inside the MR-scanner from 31 Ag–AgCl electrodes mounted on an elastic cap according to the 10–20 standard system (EasyCap-MR, Brain Products, Munich, Germany). The electrocardiogram (ECG) was registered with an additional electrode on the sternum. EEG signals were amplified with an MR-conform 32-channel amplifier (BrainAmp MR; Brain Products, Munich, Germany) that did not get saturated by MR activity. Signals were recorded at a sampling frequency of 5000 Hz and a resolution of 16 bits, referenced against FCz, using the BrainVision Recorder Software (Brain Products, Munich, Germany). The ground electrode was positioned between Fz and FPz. All impedances were kept below 5 kΩ.

Since we used a sparse imaging design with stimuli being presented in-between two consecutive volume acquisitions, gradient artifact removal from the EEG was not necessary (cf. Herrmann and Debener, 2008; Huster et al., 2012). For preprocessing, a finite impulse response (FIR) 100 Hz low-pass filter (389 points, Hamming window) and a 1.7 Hz high-pass filter (4901 points, Hann window, corresponding to a cut-off period of 1/1.7 Hz = 588 ms) was applied to the raw data. Note that filter settings were chosen such that smearing of gradient artifacts into time windows of interest were prohibited. Subsequently, filtered EEG data were down-sampled to 500 Hz and subjected to an independent components analysis (ICA) for artifact correction, using the routines provided by EEGLab (Delorme and Makeig, 2004) and *fieldtrip* (Oostenveld et al., 2011) within MATLAB 7.9 (MathWorks, Natick, MA). Note that the ECG channel was removed prior to ICA analysis. ICAs were calculated on 3-s epochs, with 1 s before and 2 s after stimulus onset. The separation of ICA components (total: 29) representing artifacts from those representing physiological EEG activity was done by visual inspection of the components' time-courses, topographies, and frequency spectra (cf. Debener et al., 2010), using custom-made fieldtrip scripts. Components either showing similar dynamics as the ECG channel or resembling electrooculogram activity as illustrated in Debener et al. (2010) were considered artifacts. Note that it has been observed that ICA-based correction of cardio-ballistic artifacts performs better than standard artifact subtraction methods (Debener et al., 2007; Jann et al., 2009). On average, 7 components were therefore excluded (range: 5–9) by using the ICA-based artifact removal within fieldtrip (Oostenveld et al., 2011).

We furthermore identified bad EEG channels after artifact removal as channels exceeding a threshold of 150 μV in more than 50% of all trials per participant. Bad channels (of which

no participant showed more than 1) were interpolated by using signal information from the average of 4–5 neighboring channels (depending on channel location).

In addition to EEG recordings inside the MR-scanner, we tested 18 different participants (9 females, mean age 25, range 20–31 years) outside the scanner. Presenting pre-recorded EPI sounds at times the scanner would have operated simulated the scanner noise. For this control group, the EEG was obtained from 64 Ag-AgCl-electrodes (58 scalp electrodes, 2 mastoids, 2 electrodes for horizontal and 2 for vertical electrooculograms) on a Brain Vision EEG system (amplifier: BrainAmp, cap: BrainCap, Brain Products, Munich, Germany), arranged according to the extended 10/20 system, (Oostenveld and Praamstra, 2001). Otherwise, stimulus presentation, EEG pre-processing and analyses were identical to the procedures described here. However, due to a technical problem with one participant, and more than 30% ICA-artifact components in two further participants, the resulting participant number of the control experiment was 15. This experiment served the purpose of testing the validity of the recordings obtained inside the scanner. Note, however, that overall magnitude differences should not be compared between the experiments inside and outside the scanner, due to different recording equipment.

ACQUISITION AND PRE-PROCESSING OF fMRI DATA

Functional MRI data were recorded with a Siemens VERIO 3.0-T MRI scanner equipped with a 12-channel head coil, while participants performed the categorization task in supine position inside the scanner. Acoustic stimuli were transmitted through MR-compatible headphones (mr confon GmbH, Magdeburg, Germany). In-ear hearing protection (Hearsafe Technologies GmbH, Cologne, Germany) reduced scanner noise by approximately 16 dB.

Seventy-five whole-brain EPI volumes (30 axial slices, thickness = 3 mm, gap = 1 mm) in each of the 4 runs were collected every 9 s (TA = 2 s; TE = 30 ms; flip angle = 90°; field of view = 192 × 192 mm; voxel size = 3 × 3× 4 mm). High-resolution, 3D MP-RAGE T1-weighted scans were used for localization and co-registration (acquired on a 3T Siemens TIM Trio scanner with a 12-channel head coil 29 months prior to the experiment, with the parameters: sagittal slices = 176, repetition time = 1300 ms, TE = 3.46 ms, flip angle = 10°, acquisition matrix = 256 × 240, voxel size = 1 × 1× 1 mm). Voxel-displacement-maps for distortion correction (Jezzard and Balaban, 1995; Hutton et al., 2002) were calculated on the basis of field maps (30 axial slices, thickness = 3 mm, gap = 1 mm, repetition time = 488 ms, TE1 = 4.92 ms, TE2 = 7.38 ms, flip angle = 60°, field of view = 192 × 192 mm, voxel size = 3 × 3× 3 mm).

Functional (T2*-weighted) and structural (T1-weighted) images were processed using Statistical Parametric Mapping (SPM8; Wellcome Department of Imaging Neuroscience, Institute of Neurology, University College of London). Functional images were first realigned using the 6-parameter affine transformation in translational (x, y, and z) and rotational (pitch, roll, and yaw) directions to reduce individual movement artifacts (Ashburner and Good, 2003). Subsequently, a mean image of each run-based image series was used to estimate unwarping

parameters, and voxel-displacement-maps were used for correcting magnetic field deformations (Jezzard and Balaban, 1995; Hutton et al., 2002). Participants' structural images were manually pre-aligned to a standardized EPI template (Ashburner and Friston, 2004) in MNI space, improving co-registration and normalization accuracy. Next, functional images were co-registered to the corresponding participants' structural images and normalized to MNI space. Functional images were then smoothed using an 8-mm full-width half-maximum Gaussian kernel and subsequently used for first-level general linear model (GLM) analyses.

ANALYSIS OF BEHAVIORAL DATA

Our behavioral dependent measures were *overall performance* and *cue utilization*. Overall performance was estimated by d', a measure of perceptual sensitivity that is independent of response bias. Perceptual sensitivity, d', was calculated from proportions of hits and false alarms according to a one-interval design (Macmillan and Creelman, 2005), where hits were defined as "category-A" responses to category-A stimuli, and false alarms were defined as "category-A" responses to category-B stimuli. Perceptual sensitivity was calculated separately for each experimental run (2 nondegraded, 2 degraded runs). In order to visualize performance over time, we additionally calculated d' values in sliding windows (size: 20 trials, step size: 1 trials), separately for the nondegraded and the degraded condition, and with the exclusion of null trials.

The measure of *cue index* quantified individual participants' cue utilization (spectral peak vs. physical duration) in the following way: First, for each condition, the likelihood of a category-A response was predicted from the stimulus' physical properties, spectral peak and duration, by means of logistic regressions. The slope of the regressions function, expressed by absolute β, indicated the degree to which the corresponding physical stimulus property influenced the categorical response ($\beta_{\text{spectral peak}}$; β_{duration}; Goudbeek et al., 2009; Scharinger et al., 2013). Note that $\beta_{\text{spectral peak}}$ and β_{duration} were estimated simultaneously. Second, the normalized difference between these β values (*cue index*) indicated participants' preference to rely on spectral peak (negative values according) or on duration (positive values).

$$\text{Cue index} = \frac{\beta_{\text{duration}} - \beta_{\text{spectral peak}}}{\beta_{\text{duration}} + \beta_{\text{spectral peak}}}$$

ANALYSIS OF EEG DATA

For the analysis of the event-related potentials (ERPs), single-trial EEG epochs were first re-referenced to linked mastoids (approximated by channels Tp9 and Tp10). Subsequently, epochs were filtered with a 20-Hz Butterworth low-pass filter and re-defined to include a pre-stimulus interval of 500 ms and a post-onset interval of 1500 ms. Baseline correction was applied by subtracting the mean amplitude of the -500 to 0 ms baseline interval from the epoch. Single-trials were averaged separately for the nondegraded and the degraded condition. Auditory N1 components (Näätänen and Picton, 1987) were identified by visual inspection in a time window between 100 and 150 ms post onset. Averaged amplitudes

for Cz within the N1 time-window were compared between conditions (nondegraded, degraded) by means of dependent-samples t-tests.

For time-frequency analyses, re-referenced EEG-data were down-sampled to 125 Hz and then decomposed with a Morlet wavelets analysis (Bertrand and Pantev, 1994), centered on windows that slid in steps of 10 ms along the temporal dimension (-1 to 2 s). In the spectral dimension, we used 1-Hz bins from 1 to 30 Hz. Wavelet widths ranged from 1 to 8 cycles, equally spaced over the 30 frequency bins. Time-frequency analyses were done separately for nondegraded and degraded trials. Mean power values of a pre-stimulus baseline interval (-500 to -50 ms) were subtracted from the epoch. A time-frequency region of interest (ROI) was chosen according to the typical alpha-band interval (7–11 Hz) and according to epochs that previously showed the suppression effect in speech (400–700 ms post onset, e.g., Obleser and Weisz, 2012; Becker et al., 2013). A consistent and symmetric posterior electrode selection for subsequent EEG/fMRI correlations was based on electrodes where alpha power was strongest in above-mentioned ROI (within the nondegraded condition). These electrodes were: CP1, CP2, P7, P3, Pz, P4, P8, POz, O1, Oz, and O2. Averaged power values in the alpha ROI was compared between conditions by means of dependent-samples t-tests.

ANALYSIS OF fMRI DATA

Activated voxels were identified using the GLM approach (Friston, 2004). At the first level, a GLM was estimated for each participant with a first-order finite impulse response (FIR; window = 2 s) and a high-pass filter with a cut-off of 128 s, representing standard settings for sparse imaging designs (cf. Peelle et al., 2010). The design matrix included regressors for *sound trials* (corresponding to volumes following sound representations), the mean-centered single-trial parametric modulator *alpha power* (obtained from the ROI defined above), and *silent trials* (corresponding to volumes following null trials). Experimental runs were included as regressors of no interest (one for each run). Six additional regressors of no-interest accounted for the realignment-induced spatial deformations of the EPI volumes.

Resulting beta-maps were restricted to gray- and white matter. This information was obtained from group-averages based on individual T1-weighted scans. On the first level, the following contrasts were calculated (separately for nondegraded and degraded conditions): *sound trials* against implicit baseline and parametric modulator *alpha power* against implicit baseline. Furthermore, we calculated the contrasts nondegraded > degraded and degraded > nondegraded.

On the second level (group level), all contrasts were compared against zero using one-sample t-tests. Additionally, for each condition (nondegraded, degraded), sound-trial contrasts (against implicit baseline) from the first level were correlated with cue index using linear regression. Differences between nondegraded and degraded conditions in Cue index/BOLD correlation were assessed by testing the slopes of the linear regressions against each other using a dependent samples t-test.

For statistical thresholding of second-level activations, we used a threshold of $p < 0.005$ combined with a cluster extent of 15

voxels that corresponds to a whole-brain significance level of $p < 0.05$, as determined from a MATLAB-implemented Monte Carlo simulation (Slotnick et al., 2003; Erb et al., 2013).

In order to visualize BOLD modulation differences across conditions, ROIs of 10 mm radii were defined using the SPM toolbox MarsBaR (Brett et al., 2002). They were centered on the peak coordinates of significant clusters identified in the whole-brain analyses. For these regions, mean regression beta values were estimated for each participant. Note that no additional tests were conducted for these regions to avoid statistical circularity. Determination of anatomical locations was based on the Automated Anatomical Labeling Atlas (AAL; Tzourio-Mazoyer et al., 2002), and PT localization followed Westbury et al. (1999).

RESULTS

BEHAVIORAL DATA

Participants performed above chance as indicated by d' values significantly greater than zero [mean $d' = 1.51$, $SD = 0.43$; $t_{(14)} = 19.19$, $p < 0.01$]. Participants' performance was characterized by a considerable improvement over the first twenty trials, as estimated from sliding-window averages of d'-values (window size: 20 trials, step size: 1 trial, **Figure 1B** top). After degradation was introduced, performance dropped to the initial level, but quickly regained a stable plateau and did not differ overall from the nondegraded condition [nondegraded vs. degraded $t_{(14)} = 1.00$, $p = 0.32$].

Cue indices marginally differed between conditions [$t_{(14)} = 1.94$, $p = 0.07$], with more negative values for the nondegraded than the degraded condition. This means that the tendency of utilizing spectral cues (i.e., a negative cue index) in the nondegraded condition decreased in the degraded condition (i.e., a positive-going cue index). However, a spectral strategy was never entirely given up, as judged from overall still negative cue indices in the degraded condition (**Figure 1B**, bottom).

EEG DATA

The N1 (100–150 ms) of the ERP showed a typical central/midline topography (inside and outside the scanner). N1 mean amplitude marginally differed between the nondegraded and the degraded condition [$t_{(14)} = 1.9$, $p = 0.08$], with more negative values in the nondegraded than in the degraded condition. This effect reached significance outside the scanner [$t_{(14)} = 7.89$, $p < 0.01$; **Figure 2A**].

Alpha power (7–11 Hz) around 400–700 ms showed a central-posterior distribution and also differed significantly between conditions, with relatively higher alpha power for the degraded than for the nondegraded condition [$t_{(14)} = 2.06$, $p = 0.04$ **Figure 2B**]. Again, this effect also held for the control experiment outside the scanner [$t_{(14)} = 2.56$, $p = 0.03$; **Figure 2C**].

In order to assess the covariation of alpha power and cue index, we calculated correlations between mean alpha power and mean cue index per participant, and in addition, separately for the nondegraded and degraded condition. Overall, mean alpha power and mean cue index did not correlate significantly [$r = 0.28$, $t_{(14)} = 1.07$, $p = 0.30$]. This held both within the nondegraded [$r = 0.23$, $t_{(14)} = 0.85$, $p = 0.41$] and the degraded condition [$r = 0.16$, $t_{(14)} = 0.60$, $p = 0.56$].

fMRI DATA

Overall auditory categorization network in parietal and temporal areas

Results from group-level whole-brain analyses showed that the categorization of nondegraded and degraded sounds (compared to baseline) lead to activations in extensive bilateral temporo-parietal clusters, with peaks in inferior parietal lobule and post-central gyrus (see **Figure 3**). Furthermore, peaks in precentral and cingulate cortex were predominantly seen for nondegraded sounds, while degraded sounds showed activations in pSTG, PT, and Heschl's gyrus. Both conditions also revealed substantial activations in middle frontal gyrus (MFG), inferior frontal gyrus (IFG), and in the dorsal medial nucleus of left Thalamus.

More activation for degraded than for nondegraded sounds was found in right IFG (extending into the insula), left and right pSTG (including parts of PT, i.e., gray matter with a likelihood of 25–45% being in PT according to Westbury et al., 1999), as well as right STG (extending into the insula). A detailed overview of the clusters is provided in **Table 2**.

Alpha power covaries with BOLD activity in pSTG, PT, and IFG

Group-level whole-brain analyses showed that single-trial alpha power correlated positively with BOLD only in the degraded condition. Here, alpha power/BOLD correlations occurred in two clusters in IFG (comprising pars triangularis and ventral orbitofrontal cortex), in one cluster located in right pSTG (with 25–45% probability of being in PT), and in one cluster in right angular gyrus. In the nondegraded condition, alpha power/BOLD correlations did not survive the statistical threshold.

Stronger modulations of BOLD by alpha power could be observed in the orbital part of right IFG, as well as in bilateral pSTG, again comprising parts of the PT (with 25–45% probability according to Westbury et al., 1999; cf. **Table 3** and **Figure 4A**).
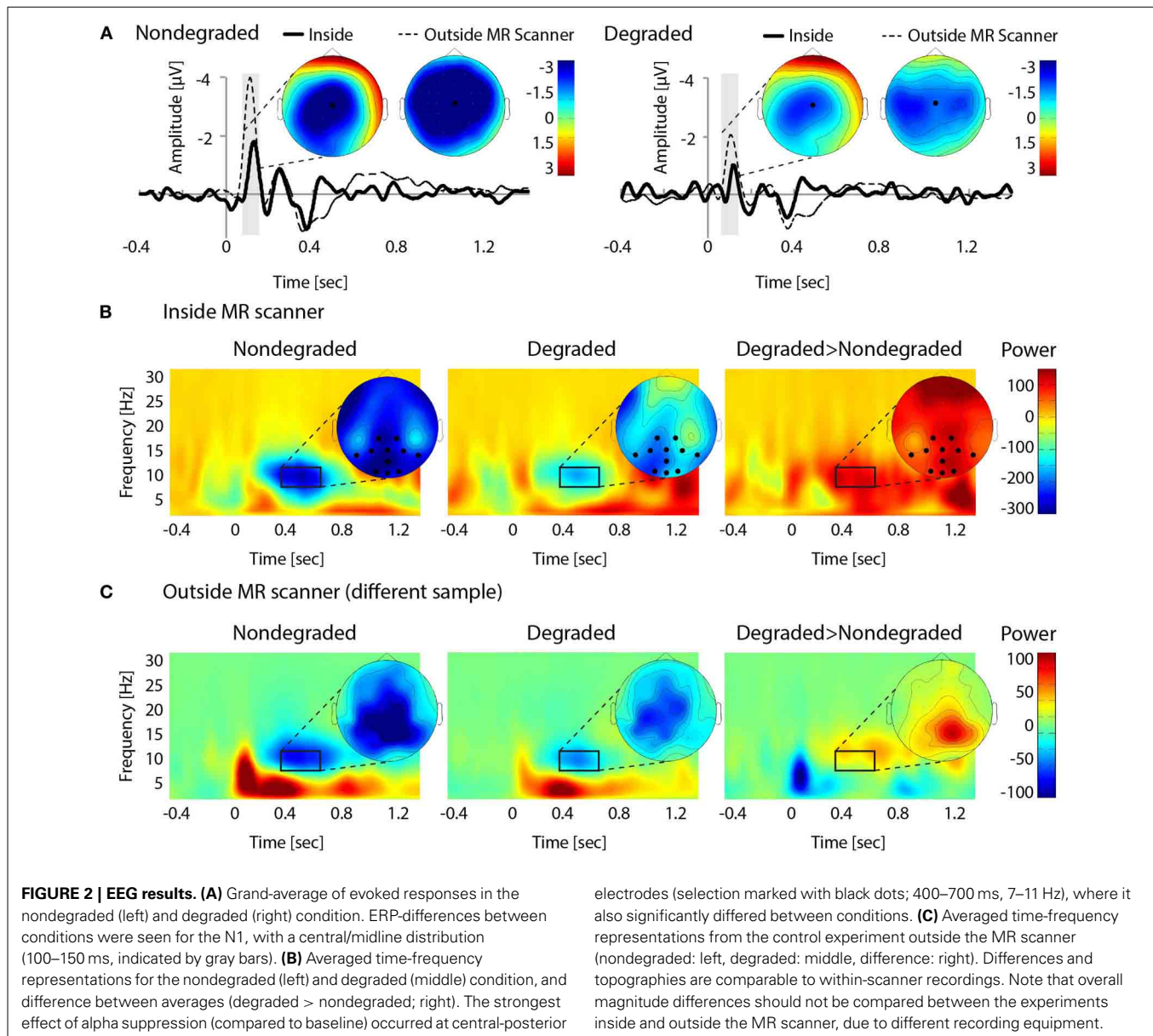
Cue index modulates BOLD activity in parietal attention and temporal auditory network

Group-level whole-brain regression analyses using the cue index showed positive correlations with BOLD in right MFG (anterior prefrontal cortex) only in the degraded condition. Here, a reduction of using spectral cues corresponded to an increased BOLD signal in anterior prefrontal cortex. By contrast, cue index/BOLD correlations in the nondegraded condition did not survive the statistical threshold.

Furthermore, positive cue index/BOLD correlations were stronger in the degraded than in the nondegraded condition in right dorso-lateral prefrontal cortex (covering parts of pars triangularis and pars opercularis), left pSTG/pSTS (extending into PT), left posterior MTG (involving parts in occipito-temporal cortex), right (ventral) IPL (involving parts of supramarginal gyrus and extending rostrally into postcentral gyrus; cf. **Table 3** and **Figure 4B**).

DISCUSSION

The two most important findings of this multimodal brain imaging study on auditory categorization are the following: First, auditory categorization of degraded stimuli yielded decreases in alpha power suppression (i.e., relative alpha power increases),

FIGURE 2 | EEG results. (A) Grand-average of evoked responses in the nondegraded (left) and degraded (right) condition. ERP-differences between conditions were seen for the N1, with a central/midline distribution (100–150 ms, indicated by gray bars). **(B)** Averaged time-frequency representations for the nondegraded (left) and degraded (middle) condition, and difference between averages (degraded > nondegraded; right). The strongest effect of alpha suppression (compared to baseline) occurred at central-posterior electrodes (selection marked with black dots; 400–700 ms, 7–11 Hz), where it also significantly differed between conditions. **(C)** Averaged time-frequency representations from the control experiment outside the MR scanner (nondegraded: left, degraded: middle, difference: right). Differences and topographies are comparable to within-scanner recordings. Note that overall magnitude differences should not be compared between the experiments inside and outside the MR scanner, due to different recording equipment.
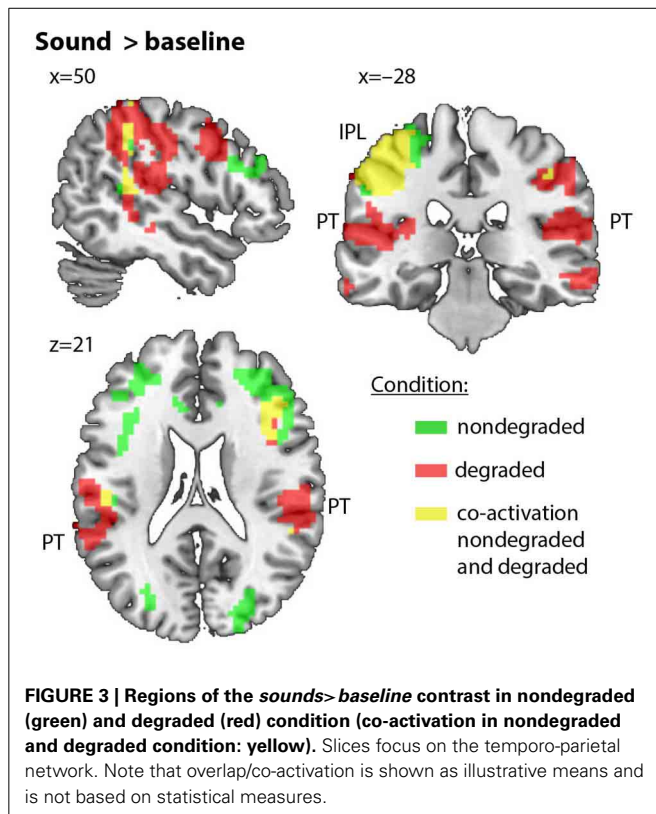
which correlated with increased activation in right PT and IFG. Second, even though the behavioral measure of cue utilization only marginally differed between conditions, less reliance on spectral cues under sound degradation corresponded to increased activation in left PT and right IPL. In the subsequent sections, these findings will be discussed in more detail.

ENHANCED ALPHA POWER DURING DEGRADED SPEECH PROCESSING

In the current study, categorizing spectrally degraded sounds was accompanied by an attenuation of alpha power suppression. That is, relatively stronger alpha power was observed for the categorization of degraded as compared to nondegraded sounds. This reduction in alpha power suppression (relative to a pre-stimulus baseline) has previously been observed in comparing spectrally degraded speech stimuli to their nondegraded (intelligible)

counter-parts (Obleser and Weisz, 2012; Becker et al., 2013). The current data thus extend previous findings by showing that increased alpha power under degradation is not restricted to speech material, but may reflect a more general process that has been interpreted before as enhanced "functional inhibition" (Jensen and Mazaheri, 2010), increased "idling" (Adrian and Matthews, 1934), or a more "active processing state" (Palva and Palva, 2011).

A parsimonious interpretation of this effect relates to the functional inhibition hypothesis of increased alpha power (e.g., Jensen and Mazaheri, 2010). According to this approach, alpha power shows a relative decrease in areas subserving the processing of to-be-attended information (Thut et al., 2006), while it increases in areas subserving the processing of to-be-ignored information (Rihs et al., 2007). Thereby, alpha power dynamics instate a

FIGURE 3 | Regions of the *sounds> baseline* contrast in nondegraded (green) and degraded (red) condition (co-activation in nondegraded and degraded condition: yellow). Slices focus on the temporo-parietal network. Note that overlap/co-activation is shown as illustrative means and is not based on statistical measures.

Table 2 | Significant clusters obtained from whole-brain analyses ($p < 0.005$, extent threshold = 15) for the contrasts sounds > baseline in each condition, and the contrast degraded sounds > nondegraded sounds.

Contrast	Area	Coordinates	Z	Extent (voxels)
Nondegraded sounds > baseline	l. IPL/BA40	−39, −13, 61	4.95	2659
	r. IFG/BA46	45, 38, 31	4.4	539
	r. IPL/SMG	42, −34, 46	4.28	470
	r. Cereb/Culmen	21, −55, −26	4.2	194
	r. Cereb/Culmen	3, −61, −32	4.18	170
	l. Thalamus	−6, −19, 7	3.87	113
	r. Cuneus	18, −91, 1	3.75	103
	l. Insula/BA13	−30, 14, 1	3.66	72
	r. Insula/BA13	30, 20, −2	3.65	61
	r. ITG/BA20	57, −46, −17	3.6	37
	l. Insula/BA13	−27, 26, −5	3.55	21
	l. Occ./BA17	−15, −91, 1	3.49	27
	l. MFG/BA10	−24, 59, −8	3.46	80
	l. pSTG/PT	−48, −46, 7	3.42	30
	r. pSTG/PT	51, −40, 13	3.17	35
Degraded sounds > baseline	l. Postcentral/IPL	−51, −22, 46	5.61	2074
	r. IPL/BA40	39, −43, 58	4.82	1563
	r. Cingulate/BA32	3, 11, 55	4.77	631
	r. Precentral/BA6	48, 5, 40	4.29	354
	l. Cuneus/BA18	−18, −100, 1	4.24	512
	r. MFG/BA11	21, 47, −11	4.24	15
	r. MFG/BA10	36, 50, 10	4	84
	r. IFG/BA47	30, 29, −2	3.7	70
	l. MFG/BA10	−33, 41, 4	3.7	79
	l. MTG/BA21	−63, −31, −14	3.64	41
	l. Thalamus	−12, −19, 10	3.47	63
	l. Insula/BA13	−30, 32, 7	3.32	75
	l. MFG/BA10	−27, 32, 25	3.2	19
	r. Cereb./Culmen	15, −52, −23	3.17	21
Degraded > Nondegraded	r. IFG/Insula	33, 14, −17	3.9	43
	l. pSTG/PT	−51, −37, 10	3.41	16
	r. STG	48, −4, −8	3.3	31
	r. pSTG/PT	54, −25, 19	3.2	30

Abbreviations are explained in the text. Coordinates are given in Montreal Neurological Institute (MNI) space.

gain mechanism for neural information processing (Jokisch and Jensen, 2007; Kerlin et al., 2010). While the functional role of alpha oscillations in auditory processing and categorization has been examined much less often and only recently (Weisz et al., 2011, 2013; Obleser and Weisz, 2012; Obleser et al., 2012; Becker et al., 2013), the interpretations provided by these previous studies are in line with the functional inhibition hypothesis. For instance, it has been observed that alpha power suppression correlates with the intelligibility of auditory (speech) input (Obleser and Weisz, 2012; Becker et al., 2013). Alpha power suppression was attenuated when auditory stimuli were degraded, that is, when comprehension was more effortful and required higher demands on attention (Obleser et al., 2012), as has been suggested for effortful listening situations before (e.g., Shinn-Cunningham and Best, 2008; Wild et al., 2012).

With respect to our data, we propose that alpha power increases gated the neural processing of acoustic information (duration vs. spectral peak) that differed in task-relevance between conditions: The introduction of spectral degradation in the second half of our experiment changed the relative informativeness or task-relevance of the spectral and duration cues, with spectral peak becoming less informative than stimulus duration. It is thus possible that enhanced alpha under degradation indexed the inhibition of spectral information processing.

Historically, however, enhanced alpha power has first been interpreted as reflecting the degree to which cortical areas are in an "idling" state (Adrian and Matthews, 1934; Niedermeyer and Silva, 2005). Consequently, reduction or suppression of alpha power was taken to index a departure from the idling mode

toward a more attentive state. While this interpretation might be applicable for the general suppression of alpha power (vs. baseline) for nondegraded and degraded conditions, it cannot explain the differences in alpha power between conditions. That is, overall performance in our experiment (and thus presumably attentional effort) was comparable between the nondegraded and degraded conditions, while alpha power increased in the latter condition. Thus, this increase in alpha power is unlikely to reflect a more pronounced idling state.

Table 3 | Significant clusters obtained from whole-brain analyses (*p* < 0.005, extent threshold = 15) for the parametric modulators alpha and cue index, together with modulation differences between conditions.

Contrast	Area	Coordinates	Z	Extent (voxels)
Alpha power by BOLD (degraded)	r. oIFG/BA47	45, 29, −8	3.37	49
	r. IFG/BA45	54, 26, 10	3.25	16
	r. pSTG/PT	51, −43, 10	3.14	31
	r. AG/BA39	36, −67, 43	3.04	16
Alpha power by BOLD (nondegraded)	–	–	*n.s.*	
Alpha power degraded > nondegraded	r. oIFG/BA47	45, 29, −11	3.32	18
	r. pSTG/PT	54, −43, 13	3	15
	l. pSTG/PT	−54, −49, 13	2.94	22
Cue index by BOLD (degraded)	r. MFG	39, 47, 4	4.46	58
Cue index by BOLD (nondegraded)	–	–	*n.s.*	
Cue index degraded > nondegraded	r. DLPFC	42, 11, 28	3.74	49
	l. pSTG/PT	−54, −40, 7	3.7	21
	r. IPL	42, −40, 40	3.53	93
	l. MTG	−45, −55, 4	3.28	22

Abbreviations are explained in the text. Coordinates are given in MNI-space.

Finally, it has been recently proposed that alpha power enhancement can also be indicative of active processing states (Palva and Palva, 2011). According to the "active processing hypothesis," enhanced alpha power underlies the coordination of neural processing in task-relevant cortical structures, particularly for higher-order attentional and executive functions. Since the participants in our experiment seemed to be reluctant to refrain from spectral cue utilization under degradation, enhanced alpha power may also relate to "listening" harder for spectral cues, i.e., to an active process of utilizing spectral cues despite their being less informative. Both the "functional inhibition" and "active processing" hypotheses can be applied to the cortical regions in which alpha power positively correlated with BOLD.

SPECTRAL DEGRADATION AND THE PLANUM TEMPORALE

In the degraded condition of our experiment, we observed positive correlations of alpha power with BOLD activations in posterior STG and PT. The posterior STG and the PT have previously been suggested to subserve the processing of spectral information, and in particular, pitch and pitch changes (Zatorre et al., 1994; Zatorre and Belin, 2001; Schönwiesner et al., 2005; Hall and Plack, 2009; Alho et al., 2014). In particular, Hall and Plack (2009) provided evidence that apart from lateral Heschl's gyrus (Schneider
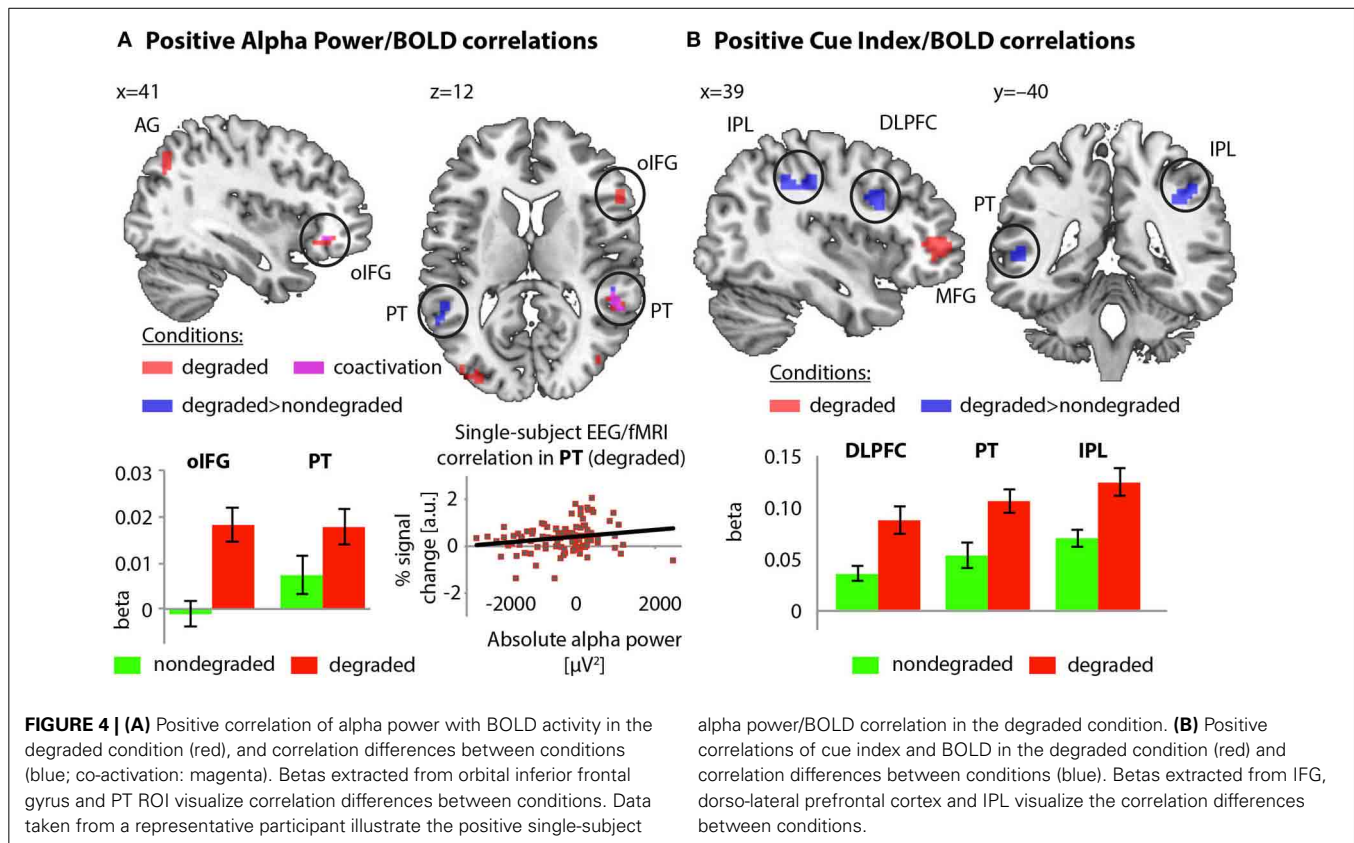
et al., 2005; Warren et al., 2005), the (right) PT supports pitch processing to a substantial degree. Importantly, Hall and Plack (2009) used stimuli that bore close resemblance to our degraded sound stimuli such that participants may have perceived and processed pitch differences between our sound categories. Altogether, the involvement of pSTG and PT in our experiment is likely to reflect spectral processing. The positive correlation of alpha power and BOLD activation in this "hub"-like structure for auditory categorization (Griffiths and Warren, 2002) can shed further light onto the relative weighting of spectral vs. duration cues under degradation.

Previous studies using simultaneous EEG-fMRI recordings have observed positive and negative correlations of alpha power with BOLD (Laufs et al., 2003; Gonçalves et al., 2006; de Munck et al., 2007; Goldman et al., 2009; Scheeringa et al., 2009, 2011; Michels et al., 2010; Liu et al., 2012). The interpretation of negative correlations of alpha power with BOLD activations follows the functional inhibition hypothesis (Foxe et al., 1998; Klimesch et al., 2007; Foxe and Snyder, 2011; Weisz et al., 2011, 2013; Klimesch, 2012; Obleser and Weisz, 2012; Obleser et al., 2012). That is, regions where activations increase with decreasing alpha power have been suggested to be relevant for attending to informative stimulus features, while regions where alpha power is positively correlated with BOLD haven been suggested to support the suppression of non-informative (task-irrelevant) stimulus features. Positive correlations of alpha power with BOLD can also be interpreted within the "active processing hypothesis" (Palva and Palva, 2011). This hypothesis relates enhanced alpha power to stronger neural coordination in cortical areas processing task-relevant information, particularly for higher-order attentional and executive functions.

Here, we observed that the posterior STG and the PT showed increased activation for degraded vs. nondegraded stimuli, and that STG and PT activations positively correlated with alpha power. This can either be interpreted with the "functional inhibition hypothesis" or the "active processing hypothesis:"

According to the "functional inhibition hypothesis," the positive correlation of alpha power with BOLD activation in (right) PT may reflect the relative inhibition of spectral information in this brain area. In detail, introduction of spectral degradation affected the informativeness of spectral peak for categorization, and corresponded to a change in cue utilization. That is, spectral peak became relatively task-irrelevant, and may have been inhibited in pSTG and PT.

According to the "active processing hypothesis," the positive correlation of alpha power and BOLD activation in pSTG and PT (particularly under degradation) may reflect the enhanced need for neural coordination in order to maintain spectral cue utilization. Overall, cue indices remained negative even after spectral information was degraded, that is, participants still relied on their initial spectral categorization strategy. For maintenance of the spectral strategy, participants might have drawn on (right) posterior STG and PT resources. Thus, the positive correlation of alpha power and BOLD in these cortical regions may index the need to listen "harder" to degraded stimulus cues that once were informative.

FIGURE 4 | (A) Positive correlation of alpha power with BOLD activity in the degraded condition (red), and correlation differences between conditions (blue; co-activation: magenta). Betas extracted from orbital inferior frontal gyrus and PT ROI visualize correlation differences between conditions. Data taken from a representative participant illustrate the positive single-subject alpha power/BOLD correlation in the degraded condition. **(B)** Positive correlations of cue index and BOLD in the degraded condition (red) and correlation differences between conditions (blue). Betas extracted from IFG, dorso-lateral prefrontal cortex and IPL visualize the correlation differences between conditions.

Finally, the "active processing hypothesis" seems to receive further support from the positive alpha power/BOLD correlations in frontal (IFG) areas. Note that Palva and Palva (2011) suggest that inhibition at lower sensory levels might be achieved by higher-level frontal functions, such that a positive alpha power/BOLD correlation in IFG may indicate that lesser reliance on spectral than on duration cues under degradation is mediated by activity in frontal regions. This may also relate to the observation that alpha power and behavioral cue utilization indices correlated only at trend-level with each other, suggesting that alpha power changes are more likely reflecting indirect, modulatory signatures of "functional inhibition" (after a stimulus while preparing a response, see also Obleser and Weisz, 2012; Wilsch et al., 2014). These signatures are dissociable from and follow in time early auditory signatures, accounting for the latency of the alpha power effect centered at around 500 ms post stimulus onset.

A ROLE OF THE RIGHT IPL IN AUDITORY ATTENTION

The behavioral tendency of disregarding spectral cues in the degraded condition of our experiment was accompanied by increased activation in anterior prefrontal cortex, and, compared to the nondegraded condition, in right IPL. In the degraded condition, right IPL showed a stronger correlation of cue index with BOLD activation than in the nondegraded condition (**Figure 4B**). As part of the fronto-parietal executive network (Posner and Dehaene, 1994; Corbetta et al., 2000), the IPL has repeatedly been found to subserve selective attention (Shaywitz et al., 2001; Behrmann et al., 2004; Salmi et al., 2009) and attentional control

(Hill and Miller, 2010). Its activation was commonly observed in situations that require flexible changes in attention during the processing of informative stimulus features or task-relevant information (Geng and Mangun, 2009; Schultz and Lennert, 2009; Gillebert et al., 2012). In line with studies supporting the IPL's role in selectively attending to the most informative stimulus feature (Jacquemot et al., 2003; Gaab et al., 2006; Husain et al., 2006; Kiefer et al., 2008; Obleser et al., 2012), changes in IPL activation might support the change in cue utilization that was necessary for successful categorization (see Henry et al., 2013 for attention to temporal features). Note however that, behaviorally, participants tried to maintain their initial strategy and overall differed only marginally in cue utilization. Therefore, this interpretation must be considered carefully and substantiated by future research.

SUMMARY

In this multi-modal imaging study, we have shown that acoustic cue utilization during auditory categorization is flexible, even though listeners seem resilient to abandon initial categorization strategies. Brain areas processing the specific acoustic information—spectral peak vs. duration—supported the change in cue preference together with areas in the fronto-parietal attention network. Our data complement previous speech-related observations of alpha power increases in adverse and effortful listening situations (Obleser and Weisz, 2012; Obleser et al., 2012; Wilsch et al., 2014). We suggest that increased alpha power under degradation mediates the relative weighting of acoustic stimulus

features. Both the "functional inhibition" and the "active processing" hypotheses can account for these findings. Importantly, the combination of behavioral, electrophysiological, and hemodynamic measures is an indispensable methodology for further investigations in auditory cognition.

ACKNOWLEDGMENTS

REFERENCES

Adrian, E. D., and Matthews, B. H. C. (1934). The interpretation of potential waves in the cortex. *J. Physiol.* 81, 440–471.

Alho, K., Rinne, T., Herron, T. J., and Woods, D. L. (2014). Stimulus-dependent activations and attention-related modulations in the auditory cortex: a meta-analysis of fMRI studies. *Hear. Res.* 307, 29–41. doi: 10.1016/j.heares.2013.08.001

Ashburner, J., and Friston, K. J. (2004). "Computational neuroanatomy," in *Human Brain Function*, eds R. S. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. Price, and S. Zeki (Amsterdam: Academic Press), 655–672.

Ashburner, J., and Good, C. D. (2003). "Spatial registration of images," in *Qualitative MRI of the Brain: Measuring Changes Caused by Disease*, ed P. Tofts (Chichester: John Wiley and Sons), 503–531. doi: 10.1002/0470869526.ch15

Becker, R., Pefkou, M., Michel, C. M., and Hervais-Adelman, A. G. (2013). Left temporal alpha-band activity reflects single word intelligibility. *Front. Syst. Neurosci.* 7:121. doi: 10.3389/fnsys.2013.00121

Behrmann, M., Geng, J. J., and Shomstein, S. (2004). Parietal cortex and attention. *Curr. Opin. Neurobiol.* 14, 212–217. doi: 10.1016/j.conb.2004.03.012

Bermudez, P., Lerch, J. P., Evans, A. C., and Zatorre, R. J. (2009). Neuroanatomical correlates of musicianship as revealed by cortical thickness and voxel-based morphometry. *Cereb. Cortex* 19, 1583–1596. doi: 10.1093/cercor/bhn196

Bertrand, O., and Pantev, C. (1994). "Stimulus frequency dependence of the transient oscillatory auditory evoked response (40 Hz) studied by electric and magnetic recordings in humans," in *Oscillatory Event-Related Brain Dynamics*, eds C. Pantev, T. Elbert, and B. Lütkenhöner (New York, NY: Plenum Press), 231–242. doi: 10.1007/978-1-4899-1307-4_17

Brett, M., Anton, J.-L., Valabregue, R., and Poline, J. B. (2002). "Region of interest analysis using an SPM toolbox," in *Paper Presented at the 8th International Conference on Functional Mapping of the Human Brain*, Sendai.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., and Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat. Neurosci.* 3, 292–297. doi: 10.1038/73009

Debener, S., Strobel, A., Sorger, B., Peters, J., Kranczioch, C., Engel, A. K., et al. (2007). Improved quality of auditory event-related potentials recorded simultaneously with 3-T fMRI: removal of the ballistocardiogram artefact. *Neuroimage* 34, 587–597. doi: 10.1016/j.neuroimage.2006.09.031

Debener, S., Thorne, J., Schneider, T. R., and Viola, F. C. (2010). "Using ICA for the analysis of multi-channel EEG data," in *Simultaneous EEG and fMRI: Recording, Analysis, and Application*, eds M. Ullsperger and S. Debener (Oxford: Oxford University Press), 121–134. doi: 10.1093/acprof:oso/9780195372731.003.0008

Delorme, A., and Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21. doi: 10.1016/j.jneumeth.2003.10.009

de Munck, J. C., Gonçalves, S. I., Huijboom, L., Kuijer, J. P. A., Pouwels, P. J. W., Heethaar, R. M., et al. (2007). The hemodynamic response of the alpha rhythm: an EEG/fMRI study. *Neuroimage* 35, 1142–1151. doi: 10.1016/j.neuroimage.2007.01.022

Desai, R., Liebenthal, E., Waldron, E., and Binder, J. R. (2008). Left posterior temporal regions are sensitive to auditory categorization. *J. Cogn. Neurosci.* 20, 1174–1188. doi: 10.1162/jocn.2008.20081

Drullman, R., Festen, J. M., and Plomp, R. (1994). Effect of temporal envelope smearing on speech reception. *J. Acoust. Soc. Am.* 95, 1053–1064. doi: 10.1121/1.408467

Erb, J., Henry, M. J., Eisner, F., and Obleser, J. (2012). Auditory skills and brain morphology predict individual differences in adaptation to degraded speech. *Neuropsychologia* 50, 2154–2164. doi: 10.1016/j.neuropsychologia.2012.05.013

Erb, J., Henry, M. J., Eisner, F., and Obleser, J. (2013). The brain dynamics of rapid perceptual adaptation to adverse listening conditions. *J. Neurosci.* 33, 10688–10697. doi: 10.1523/JNEUROSCI.4596-12.2013

Foxe, J. J., Simpson, G. V., and Ahlfors, S. P. (1998). Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. *Neuroreport* 9, 3929–3933. doi: 10.1097/00001756-199812010-00030

Foxe, J. J., and Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Front. Percept. Sci.* 2:154. doi: 10.3389/fpsyg.2011.00154

Friston, K. J. (2004). "Experimental design and statistical parametric mapping," in *Human Brain Function*, eds R. S. Frackowiak, K. J. Friston, C. D. Frith, R. J. Dolan, C. Price, and S. Zeki (Amsterdam: Academic Press), 599–632.

Gaab, N., Gaser, C., and Schlaug, G. (2006). Improvement-related functional plasticity following pitch memory training. *Neuroimage* 31, 255–263. doi: 10.1016/j.neuroimage.2005.11.046

Geng, J. J., and Mangun, G. R. (2009). Anterior intraparietal sulcus is sensitive to bottom-up attention driven by stimulus salience. *J. Cogn. Neurosci.* 21, 1584–1601. doi: 10.1162/jocn.2009.21103

Gillebert, C. R., Dyrholm, M., Vangkilde, S., Kyllingsbæk, S., Peeters, R., and Vandenberghe, R. (2012). Attentional priorities and access to short-term memory: parietal interactions. *Neuroimage* 62, 1551–1562. doi: 10.1016/j.neuroimage.2012.05.038

Glasberg, B. R., and Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hear. Res.* 47, 103–138. doi: 10.1016/0378-5955(90)90170-T

Goldman, R. I., Stern, J. M., Engel, J. Jr., and Cohen, M. S. (2002). Simultaneous EEG and fMRI of the alpha rhythm. *Neuroreport* 13, 2487–2492. doi: 10.1097/00001756-200212200-00022

Goldman, R. I., Wei, C.-Y., Philiastides, M. G., Gerson, A. D., Friedman, D., Brown, T. R., et al. (2009). Single-trial discrimination for integrating simultaneous EEG and fMRI: identifying cortical areas contributing to trial-to-trial variability in the auditory oddball task. *Neuroimage* 47, 136–147. doi: 10.1016/j.neuroimage.2009.03.062

Gonçalves, S. I., de Munck, J. C., Pouwels, P. J. W., Schoonhoven, R., Kuijer, J. P. A., Maurits, N. M., et al. (2006). Correlating the alpha rhythm to BOLD using simultaneous EEG/fMRI: inter-subject variability. *Neuroimage* 30, 203–213. doi: 10.1016/j.neuroimage.2005.09.062

Goudbeek, M., Swingley, D., and Smits, R. (2009). Supervised and unsupervised learning of multidimensional acoustic categories. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1913–1933. doi: 10.1037/a0015781

Griffiths, T. D., and Warren, J. D. (2002). The planum temporale as a computational hub. *Trends Neurosci.* 25, 348–353. doi: 10.1016/S0166-2236(02)02191-4

Guenther, F. H., Nieto-Castanon, A., Ghosh, S. S., and Tourville, J. A. (2004). Representation of sound categories in auditory cortical maps. *J. Speech Lang. Hear. Res.* 47, 46–57. doi: 10.1044/1092-4388(2004/005)

Hall, D. A., Haggard, M. P., Akeroyd, M. A., Palmer, A. R., Summerfield, A. Q., Elliott, M. R., et al. (1999). "Sparse temporal sampling" in auditory fMRI. *Hum. Brain Mapp.* 7, 213–223.

Hall, D. A., Johnsrude, I. S., Haggard, M. P., Palmer, A. R., Akeroyd, M. A., and Summerfield, A. Q. (2002). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 12, 140–149. doi: 10.1093/cercor/12.2.140

Hall, D. A., and Plack, C. J. (2009). Pitch processing sites in the human auditory brain. *Cereb. Cortex* 19, 576–585. doi: 10.1093/cercor/bhn108

Henry, M. J., Herrmann, B., and Obleser, J. (2013). Selective attention to temporal features on nested time scales. *Cereb. Cortex*. doi: 10.1093/cercor/bht240. [Epub ahead of print].

Herrmann, C. S., and Debener, S. (2008). Simultaneous recording of EEG and BOLD responses: a historical perspective. *Int. J. Psychophysiol.* 67, 161–168. doi: 10.1016/j.ijpsycho.2007.06.006

Hill, K. T., and Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cereb. Cortex* 20, 583–590. doi: 10.1093/cercor/bhp124

Holt, L. L., and Lotto, A. J. (2006). Cue weighting in auditory categorization: implications for first and second language acquisition. *J. Acoust. Soc. Am.* 119, 3059–3071. doi: 10.1121/1.2188377

Husain, F. T., Fromm, S. J., Pursley, R. H., Hosey, L. A., Braun, A. R., and Horwitz, B. (2006). Neural bases of categorization of simple speech and nonspeech sounds. *Hum. Brain Mapp.* 27, 636–651. doi: 10.1002/hbm.20207

Huster, R. J., Debener, S., Eichele, T., and Herrmann, C. S. (2012). Methods for simultaneous EEG-fMRI: an introductory review. *J. Neurosci.* 32, 6053–6060. doi: 10.1523/JNEUROSCI.0447-12.2012

Hutton, C., Bork, A., Josephs, O., Deichmann, R., Ashburner, J., and Turner, R. (2002). Image distortion correction in fMRI: a quantitative evaluation. *Neuroimage* 16, 217–240. doi: 10.1006/nimg.2001.1054

Jacquemot, C., Pallier, C., LeBihan, D., Dehaene, S., and Dupoux, E. (2003). Phonological grammar shapes the auditory cortex: a functional magnetic resonance imaging study. *J. Neurosci.* 23, 9541–9546.

Jann, K., Dierks, T., Boesch, C., Kottlow, M., Strik, W., and Koenig, T. (2009). BOLD correlates of EEG alpha phase-locking and the fMRI default mode network. *Neuroimage* 45, 903–916. doi: 10.1016/j.neuroimage.2009.01.001

Jensen, O., and Mazaheri, A. (2010). Shaping functional architecture by oscillatory alpha activity: gating by inhibition. *Front. Hum. Neurosci.* 4:186. doi: 10.3389/fnhum.2010.00186

Jezzard, P., and Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magn. Reson. Med.* 34, 65–73. doi: 10.1002/mrm.1910340111

Jokisch, D., and Jensen, O. (2007). Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. *J. Neurosci.* 27, 3244–3251. doi: 10.1523/JNEUROSCI.5399-06.2007

Kerlin, J. R., Shahin, A. J., and Miller, L. M. (2010). Attentional gain control of ongoing cortical speech representations in a "cocktail party." *J. Neurosci.* 30, 620–628. doi: 10.1523/JNEUROSCI.3631-09.2010

Kiefer, M., Sim, E.-J., Herrnberger, B., Grothe, J., and Hoenig, K. (2008). The sound of concepts: four markers for a link between auditory and conceptual brain systems. *J. Neurosci.* 28, 12224–12230. doi: 10.1523/JNEUROSCI.3579-08.2008

Klimesch, W. (2012). Alpha-band oscillations, attention, and controlled access to stored information. *Trends Cogn. Sci.* 16, 606–617. doi: 10.1016/j.tics.2012.10.007

Klimesch, W., Sauseng, P., and Hanslmayr, S. (2007). EEG alpha oscillations: the inhibition-timing hypothesis. *Brain Res. Rev.* 53, 63–88. doi: 10.1016/j.brainresrev.2006.06.003

Laufs, H., Kleinschmidt, A., Beyerle, A., Eger, E., Salek-Haddadi, A., Preibisch, C., et al. (2003). EEG-correlated fMRI of human alpha activity. *Neuroimage* 19, 1463–1476. doi: 10.1016/S1053-8119(03)00286-6

Liu, Z., de Zwart, J. A., Yao, B., van Gelderen, P., Kuo, L.-W., and Duyn, J. H. (2012). Finding thalamic BOLD correlates to posterior alpha EEG. *Neuroimage* 63, 1060–1069. doi: 10.1016/j.neuroimage.2012.08.025

Macmillan, N. A., and Creelman, C. D. (2005). *Detection Theory: A User's Guide.* Mahwah, NJ: Erlbaum.

Michels, L., Bucher, K., Lüchinger, R., Klaver, P., Martin, E., Jeanmonod, D., et al. (2010). Simultaneous EEG-fMRI during a working memory task: modulations in low and high frequency bands. *PLoS ONE* 5:e10298. doi: 10.1371/journal.pone.0010298

Moosmann, M., Ritter, P., Krastel, I., Brink, A., Thees, S., Blankenburg, F., et al. (2003). Correlates of alpha rhythm in functional magnetic resonance imaging and near infrared spectroscopy. *Neuroimage* 20, 145–158. doi: 10.1016/S1053-8119(03)00344-6

Näätänen, R., and Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425. doi: 10.1111/j.1469-8986.1987.tb00311.x

Nahum, M., Nelken, I., and Ahissar, M. (2008). Low-level information and high-level perception: the case of speech in noise. *PLoS Biol.* 6:e216. doi: 10.1371/journal.pbio.0060126

Niedermeyer, E., and Silva, F. H. L. D. (2005). *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields.* Philadelphia, PA: Lippincott Williams and Wilkins.

Obleser, J., and Eisner, F. (2009). Pre-lexical abstraction of speech in the auditory cortex. *Trends Cogn. Sci.* 13, 14–19. doi: 10.1016/j.tics.2008.09.005

Obleser, J., and Kotz, S. A. (2010). Expectancy constraints in degraded speech modulate the language comprehension network. *Cereb. Cortex* 20, 633–640. doi: 10.1093/cercor/bhp128

Obleser, J., and Weisz, N. (2012). Suppressed alpha oscillations predict intelligibility of speech and its acoustic details. *Cereb. Cortex* 22, 2466–2477. doi: 10.1093/cercor/bhr325

Obleser, J., Wöstmann, M., Hellbernd, N., Wilsch, A., and Maess, B. (2012). Adverse listening conditions and memory load drive a common alpha oscillatory network. *J. Neurosci.* 32, 12376–12383. doi: 10.1523/JNEUROSCI.4908-11.2012

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869

Oostenveld, R., and Praamstra, P. (2001). The five percent electrode system for high-resolution EEG and ERP measurements. *Clin. Neurophysiol.* 112, 713–719. doi: 10.1016/S1388-2457(00)00527-7

Palva, S., and Palva, J.M. (2011). Functional roles of alpha-band phase synchronization in local and large-scale cortical networks. *Front. Psychol.* 2:204. doi: 10.3389/fpsyg.2011.00204

Peelle, J. E., Eason, R. J., Schmitter, S., Schwarzbauer, C., and Davis, M. H. (2010). Evaluating an acoustically quiet EPI sequence for use in fMRI studies of speech and auditory processing. *Neuroimage* 52, 1410–1419. doi: 10.1016/j.neuroimage.2010.05.015

Poissant, S. F., Whitmal, N. A. 3rd., and Freyman, R. L. (2006). Effects of reverberation and masking on speech intelligibility in cochlear implant simulations. *J. Acoust. Soc. Am.* 119, 1606–1615. doi: 10.1121/1.2168428

Posner, M. I., and Dehaene, S. (1994). Attentional networks. *Trends Neurosci.* 17, 75–79. doi: 10.1016/0166-2236(94)90078-7

Rihs, T. A., Michel, C. M., and Thut, G. (2007). Mechanisms of selective inhibition in visual spatial attention are indexed by alpha-band EEG synchronization. *Eur. J. Neurosci.* 25, 603–610. doi: 10.1111/j.1460-9568.2007.05278.x

Rinne, T., Stecker, G. C., Kang, X., Yund, E. W., Herron, T. J., and Woods, D. L. (2007). Attention modulates sound processing in human auditory cortex but not the inferior colliculus. *Neuroreport* 18, 1311–1314. doi: 10.1097/WNR.0b013e32826fb3bb

Ritter, P., and Villringer, A. (2006). Simultaneous EEG-fMRI. *Neurosci. Biobehav. Rev.* 30, 823–838. doi: 10.1016/j.neubiorev.2006.06.008

Ronnberg, J., Rudner, M., Foo, C., and Lunner, T. (2008). Cognition counts: a working memory system for ease of language understanding (ELU). *Int. J. Audiol.* 47, S99–S105. doi: 10.1080/14992020802301167

Rosen, S., Faulkner, A., and Wilkinson, L. (1999). Adaptation by normal listeners to upward spectral shifts of speech: implications for cochlear implants. *J. Acoust. Soc. Am.* 106, 3629–3636. doi: 10.1121/1.428215

Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A.-L., D'Esposito, M., et al. (2012). Alpha-band phase synchrony is related to activity in the fronto-parietal adaptive control network. *J. Neurosci.* 32, 14305–14310. doi: 10.1523/JNEUROSCI.1358-12.2012

Sadaghiani, S., Scheeringa, R., Lehongre, K., Morillon, B., Giraud, A.-L., and Kleinschmidt, A. (2010). Intrinsic connectivity networks, alpha oscillations, and tonic alertness: a simultaneous electroencephalography/functional magnetic resonance imaging study. *J. Neurosci.* 30, 10243–10250. doi: 10.1523/JNEUROSCI.1004-10.2010

Salmi, J., Rinne, T., Koistinen, S., Salonen, O., and Alho, K. (2009). Brain networks of bottom-up triggered and top-down controlled shifting of auditory attention. *Brain Res.* 1286, 155–164. doi: 10.1016/j.brainres.2009.06.083

Scharinger, M., Henry, M. J., Erb, J., Meyer, L., and Obleser, J. (2014). Thalamic and parietal brain morphology predicts auditory category learning. *Neuropsychologia* 53, 75–83. doi: 10.1016/j.neuropsychologia.2013.09.012

Scharinger, M., Henry, M. J., and Obleser, J. (2013). Prior experience with negative spectral correlations promotes information integration during auditory category learning. *Mem. Cogn.* 41, 752–768. doi: 10.3758/s13421-013-0294-9

Scheeringa, R., Fries, P., Petersson, K.-M., Oostenveld, R., Grothe, I., Norris, D. G., et al. (2011). Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. *Neuron* 69, 572–583. doi: 10.1016/j.neuron.2010.11.044

Scheeringa, R., Petersson, K. M., Kleinschmidt, A., Jensen, O., and Bastiaansen, M. C. M. (2012). EEG alpha power modulation of fMRI resting-state connectivity. *Brain Connect.* 2, 254–264. doi: 10.1089/brain.2012.0088

Scheeringa, R., Petersson, K. M., Oostenveld, R., Norris, D. G., Hagoort, P., and Bastiaansen, M. C. M. (2009). Trial-by-trial coupling between EEG and BOLD identifies networks related to alpha and theta EEG power increases during working memory maintenance. *Neuroimage* 44, 1224–1238. doi: 10.1016/j.neuroimage.2008.08.041

Schneider, P., Sluming, V., Roberts, N., Scherg, M., Goebel, R., Specht, H. J., et al. (2005). Structural and functional asymmetry of lateral Heschl's gyrus reflects pitch perception preference. *Nat. Neurosci.* 8, 1241–1247. doi: 10.1038/nn1530

Schönwiesner, M., Rübsamen, R., and von Cramon, D. Y. (2005). Hemispheric asymmetry for spectral and temporal processing in the human antero-lateral auditory belt cortex. *Eur. J. Neurosci.* 22, 1521–1528. doi: 10.1111/j.1460-9568.2005.04315.x

Schultz, J., and Lennert, T. (2009). BOLD signal in intraparietal sulcus covaries with magnitude of implicitly driven attention shifts. *Neuroimage* 45, 1314–1328. doi: 10.1016/j.neuroimage.2009.01.012

Scott, S. K., Rosen, S., Lang, H., and Wise, R. J. S. (2006). Neural correlates of intelligibility in speech investigated with noise vocoded speech–a positron emission tomography study. *J. Acoust. Soc. Am.* 120, 1075–1083. doi: 10.1121/1.2216725

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science* 270, 303–304. doi: 10.1126/science.270.5234.303

Sharda, M., and Singh, N. C. (2012). Auditory perception of natural sound categories - An fMRI study. *Neuroscience* 214, 49–58. doi: 10.1016/j.neuroscience.2012.03.053

Shaywitz, B. A., Shaywitz, S. E., Pugh, K. R., Fulbright, R. K., Skudlarski, P., Mencl, W. E., et al. (2001). The functional neural architecture of components of attention in language-processing tasks. *Neuroimage* 13, 601–612. doi: 10.1006/nimg.2000.0726

Shinn-Cunningham, B. G., and Best, V. (2008). Selective attention in normal and impaired hearing. *Trends Amplif.* 12, 283–299. doi: 10.1177/1084713808325306

Slotnick, S. D., Moo, L. R., Segal, J. B., and Hart, J. Jr. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Brain Res. Cogn. Brain Res.* 17, 75–82. doi: 10.1016/S0926-6410(03)00082-X

Smits, R., Sereno, J., and Jongman, A. (2006). Categorization of sounds. *J. Exp. Psychol. Hum. Percept. Perform.* 32, 733–754. doi: 10.1037/0096-1523.32.3.733

Thut, G., Nietzel, A., Brandt, S. A., and Pascual-Leone, A. (2006). Alpha-band electroencephalographic activity over occipital cortex indexes visuospatial attention bias and predicts visual target detection. *J. Neurosci.* 26, 9494–9502. doi: 10.1523/JNEUROSCI.0875-06.2006

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15, 273–289. doi: 10.1006/nimg.2001.0978

Warren, J. D., Jennings, A. R., and Griffiths, T. D. (2005). Analysis of the spectral envelope of sounds by the human brain. *Neuroimage* 24, 1052–1057. doi: 10.1016/j.neuroimage.2004.10.031

Weissman, D. H., Warner, L. M., and Woldorff, M. G. (2009). Momentary reductions of attention permit greater processing of irrelevant stimuli. *Neuroimage* 48, 609–615. doi: 10.1016/j.neuroimage.2009.06.081

Weisz, N., Hartmann, T., Müller, N., Lorenz, I., and Obleser, J. (2011). Alpha rhythms in audition: cognitive and clinical perspectives. *Front. Psychol.* 2:73. doi: 10.3389/fpsyg.2011.00073

Weisz, N., Müller, N., Jatzev, S., and Bertrand, O. (2013). Oscillatory alpha modulations in right auditory regions reflect the validity of acoustic cues in an auditory spatial attention task. *Cereb. Cortex.* doi: 10.1093/cercor/bht113. [Epub ahead of print].

Westbury, C. F., Zatorre, R. J., and Evans, A. C. (1999). Quantifying variability in the planum temporale: a probability map. *Cereb. Cortex* 9, 392–405. doi: 10.1093/cercor/9.4.392

Wild, C. J., Yusuf, A., Wilson, D. E., Peelle, J. E., Davis, M. H., and Johnsrude, I. S. (2012). Effortful listening: the processing of degraded speech depends critically on attention. *J. Neurosci.* 32, 14010–14021. doi: 10.1523/JNEUROSCI.1528-12.2012

Wilsch, A., Henry, M. J., Herrmann, B., Maess, B., and Obleser, J. (2014). Alpha oscillatory dynamics index temporal expectation benefits in working memory. *Cereb. Cortex.* doi: 10.1093/cercor/bhu1004. [Epub ahead of print].

Yantis, S. (1993). Stimulus-driven attentional capture and attentional control settings. *J. Exp. Psychol. Hum. Percept. Perform.* 19, 676–681. doi: 10.1037/0096-1523.19.3.676

Yantis, S. (2008). The neural basis of selective attention: cortical sources and targets of attentional modulation. *Curr. Dir. Psychol. Sci.* 17, 86–90. doi: 10.1111/j.1467-8721.2008.00554.x

Zatorre, R. J., and Belin, P. (2001). Spectral and temporal processing in human auditory cortex. *Cereb. Cortex* 11, 946–953. doi: 10.1093/cercor/11.10.946

Zatorre, R. J., Evans, A. C., and Meyer, E. (1994). Neural mechanisms underlying melodic perception and memory for pitch. *J. Neurosci.* 14, 1908–1919.

Zion Golumbic, E. M., Ding, N., Bickel, S., Lakatos, P., Schevon, C. A., McKhann, G. M., et al. (2013). Mechanisms underlying selective neuronal tracking of attended speech at a "cocktail party." *Neuron* 77, 980–991. doi: 10.1016/j.neuron.2012.12.037

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.